

# **Reform Support Network**

---

## **Great Teachers and Leaders: State Considerations on Building Systems of Educator Effectiveness**

---

**Spring 2011**

This document was produced under U.S. Department of Education Contract No. ED-ESE-10-0-0087 with ICF International. The views expressed herein do not necessarily represent the positions or policies of the Department of Education. No official endorsement by the U.S. Department of Education of any product, commodity, service or enterprise mentioned in this publication is intended or should be inferred.

## Introduction

On this point education research is clear – effective instruction matters. Teachers are the single most important school-level influence on student achievement.<sup>1</sup> So it is no surprise that, with 43 states and the District of Columbia adopting college- and career-ready Common Core State Standards, and 45 states and the District of Columbia involved in Race to the Top assessment consortia, states and districts are looking to ensure that they have a workforce that can deliver on rigorous student performance expectations. Recently, there has been an unprecedented focus across the nation on developing *systems of educator effectiveness* – cultivating highly-effective teachers and leaders by reexamining and realigning a range of policies and practices for recruiting, developing, retaining, and rewarding teachers and principals.

As part of this focus on systems of educator effectiveness, states and districts are rethinking the ways they *evaluate* teachers by improving the processes and the tools they use for assessing teachers, in particular by making *student* performance a significant criterion among multiple measures of *teacher* effectiveness.

While *performance-based teacher evaluations* are the focus of this paper, there are other crucial pieces to a full-fledged system of educator effectiveness. Evaluation of principals is a critical component. Professional development that is tailored to address the particular needs of individual teachers and principals, including those identified through performance-based evaluations, is also important. Some experts suggest it is critical that policies for promotion, tenure, compensation, and dismissal also be connected to performance-based evaluations. Teacher and leader recruitment policies, the structure and content of teacher and principal preparation programs, and the requirements for entry into the profession also have the potential to shape an overall system of educator effectiveness.

Why start with a focus on teacher evaluations as part of a system of educator effectiveness? Many experts argue that performance-based teacher evaluations – evaluations that include student achievement results as a significant measure of teacher performance, and also include meaningful, regular observations of classroom practice, and timely and detailed feedback to teachers – are an important foundation for a comprehensive and coherent system of educator effectiveness that aims to raise student achievement. Recent research by The New Teacher Project suggests that today, teacher effectiveness “is not measured, recorded, or used to inform decision-making in any meaningful way.” Their report, *The Widget Effect*, found that across the nation, teacher evaluations fail to differentiate performance. The result is that current teacher evaluations provide little data or information that could be used to give teachers the training and tools they need to be effective, better identify and meet individual professional development

---

<sup>1</sup> Hanushek, Eric A., and Steve G. Rivkin. 2010. "Generalizations about Using Value-Added Measures of Teacher Quality." *American Economic Review* 100(2):267–71; Rockoff, Jonah. 2003. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data," Harvard University. Sanders, W. L., & Horn, S. P. (1994). The Tennessee Value-Added Assessment System (TVAAS) Mixed model methodology in educational assessment. *Journal of Personnel Evaluation in Education*, 8(1), 299-311. Wenglinsky, H. (2000, October). How teaching matters: Bringing the classroom back into discussions of teacher quality. Princeton, NJ: The Milken Family Foundation and Educational Testing Service.

needs, provide targeted intervention to help struggling teachers, or reward the accomplishments of effective teachers.

The report that follows is an effort by the Reform Support Network to broadly share the key questions, themes, and challenges related specifically to the development of performance-based teacher evaluation systems discussed by technical experts and Race to the Top state grantees during the first six months of the Reform Support Network's Teacher and Leader Effectiveness Community of Practice. The group has explored a number of issues critical to states making sound initial choices about overarching teacher evaluation systems design, student growth models, measuring performance in non-tested grades and subjects, and teacher observation instruments.

---

### **Some Design Principles for Effective Teacher Evaluation Systems**

- 1. All teachers should be evaluated annually.*
- 2. Evaluations should be based on clear standards of instructional excellence that prioritize student learning.*
- 3. Evaluations should consider multiple measures, with emphasis on a teacher's impact on student academic growth.*
- 4. Evaluations should employ four to five rating levels.*
- 5. Evaluations should encourage frequent observations and constructive critical feedback.*
- 6. Evaluation outcomes must matter; evaluation data should be a major factor in key employment decisions.*

**-The New Teacher Project**

---

## **Background**

### **Race to the Top**

The \$4.35 billion Race to the Top Fund represents an unprecedented federal investment in reform. The initial grants are supporting eleven states and the District of Columbia in their efforts to implement comprehensive, coherent, statewide education reform across four key areas:

- Adopting standards and assessments that prepare students to succeed in college and the workplace;

- Building data systems that measure student growth and success, and inform teachers and principals how to improve instruction;
- Recruiting, developing, rewarding, and retaining effective teachers and principals, especially where they are needed most; and
- Turning around their lowest-performing schools.

The Race to the Top program has fundamentally redefined the education landscape in America by providing resources to states to lead comprehensive reform. A total of 46 states and the District of Columbia submitted bold, comprehensive Race to the Top plans; of the 35 states that applied and did not receive funding, many are still moving forward with those state plans.<sup>2</sup>

When it comes to great teachers and leaders, the 12 Race to the Top grantees are working to develop comprehensive systems of educator effectiveness by adopting clear approaches to measuring student growth; designing and implementing rigorous, transparent, and fair evaluation systems for teachers and principals; conducting annual evaluations that include timely and constructive feedback; and using evaluation information to inform professional development, compensation, promotion, retention, and tenure decisions.

Race to the Top states are not the only states tackling these issues. A 2010 review of state teacher policies by the National Council on Teacher Quality (NCTQ) finds that 21 states are now requiring annual evaluations of all teachers and 16 states are requiring that student achievement be incorporated into teacher evaluations.<sup>3</sup>

As states across the nation continue their focus on increasing effectiveness, the Reform Support Network is committed to making information about these efforts – from all states – widely available as a means of helping states to offer mutual support, lessons learned, expertise, and resources to aid one another on the road to reform.

---

**Through Race to the Top, states are working to:**

*“design and implement rigorous, transparent, and fair evaluation systems for teachers and principals that differentiate effectiveness using multiple rating categories that take into account data on student growth...as a significant factor.”*

**For more information on Race to the Top see:**

<http://www2.ed.gov/programs/racetothetop>

---

---

<sup>2</sup> See “Race to the Top Has Unique Role to Play in Reforming Schools for the Future” at <http://www.ed.gov/blog/2010/09/race-to-the-top-has-unique-role-to-play-in-reforming-schools-for-the-future>.

<sup>3</sup> See National Council on Teacher Quality. 2010. *Blueprint for Change: National Summary* at [www.nctq.org/stpy](http://www.nctq.org/stpy).

---

### **State Plans for Improving Teacher and Principal Effectiveness Based on Performance**

State Race to the Top applications were evaluated on the extent to which each state, in collaboration with its participating local educational agencies (LEAs), has a high-quality plan with ambitious yet achievable annual targets to ensure that participating LEAs:

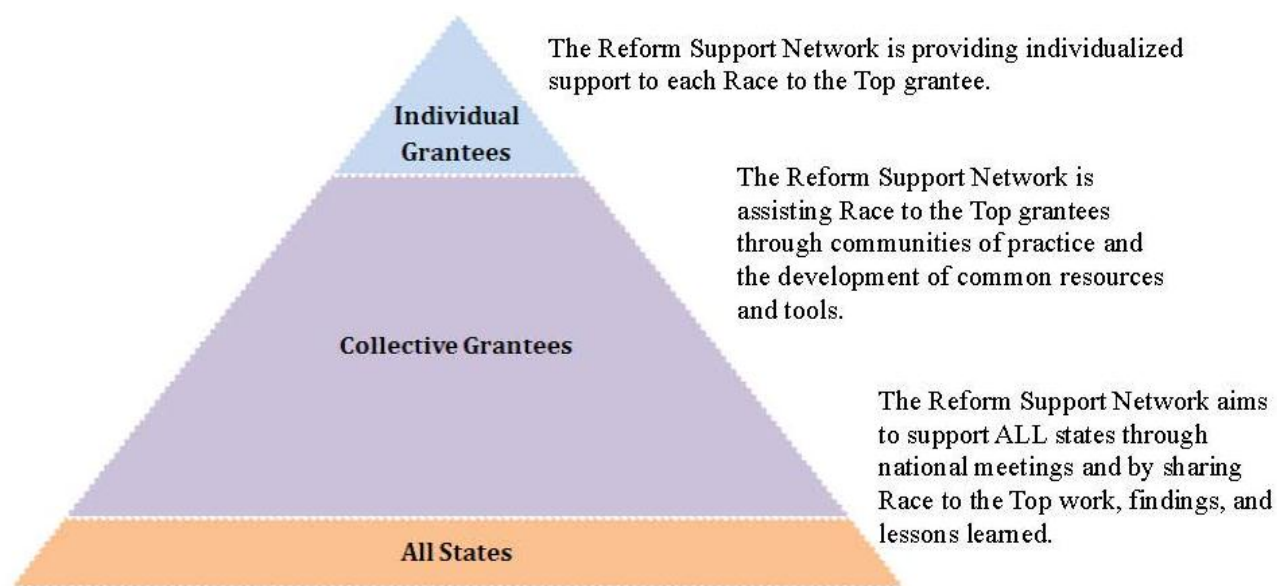
- Establish clear approaches to measuring student growth and measure it for each individual student;
- Design and implement rigorous, transparent, and fair evaluation systems for teachers and principals that (a) differentiate effectiveness using multiple rating categories that include data on student growth as a significant factor, and (b) are designed and developed with teacher and principal involvement;
- Conduct annual evaluations of teachers and principals that include timely and constructive feedback; as part of such evaluations, provide teachers and principals with data on student growth for their students, classes, and school; and
- Use these evaluations, at a minimum, to inform decisions regarding (a) developing teachers and principals, including by providing relevant coaching, induction support, and/or professional development; (b) compensating, promoting, and retaining teachers and principals, including by providing opportunities for highly effective teachers and principals to obtain additional compensation and be given additional responsibilities; (c) whether to grant tenure and/or full certification (where applicable) to teachers and principals using rigorous standards and streamlined, transparent, and fair procedures; and (d) removing ineffective tenured and untenured teachers and principals after they have had ample opportunities to improve, and ensuring that such decisions are made using rigorous standards and streamlined, transparent, and fair procedures.

**- Race to the Top application criteria**

---

## **Reform Support Network**

The Reform Support Network is funded by the U.S. Department of Education to assist Race to the Top grantee states in implementing their comprehensive education reform plans. The Network is also committed to supporting all reform-minded states by widely sharing information on the kinds of education policies being adopted as part of Race to the Top.



The Reform Support Network's goal is to support Race to the Top by:

- Building capacity to execute and sustain reforms and continuously improve outcomes;
- Providing technical assistance to Race to the Top states;
- Facilitating collaboration across states; and
- Identifying and sharing promising and effective practices across states.

## Communities of Practice

*Communities of practice are “groups of people who share a concern, a set of problems, or passion about a topic, and who deepen their knowledge and expertise in this area by interacting on an ongoing basis.”*

**-Wenger and Snyder**

---

### Teacher and Leader Effectiveness Community of Practice

As part of an effort to provide cross-grantee support, one of the strategies the Reform Support Network has adopted is to establish communities of practice (CoPs) to provide grantees with opportunities for cross-state learning and peer collaboration as well as to provide support to states from experts in the field. The purpose of the CoPs is to enhance state capacity for implementing reforms by providing for peer-to-peer learning, expert advice, model sharing, and collaboration on common needs.

In the fall of 2010, the Reform Support Network launched a Teacher and Leader Effectiveness Community of Practice to support RTT grantees in developing and implementing systems of educator effectiveness.

Initial work in this CoP has included collaborating to:

- Examine and develop practical approaches to measuring student learning using value-added models and student growth measures for the purpose of evaluating teacher performance;
- Explore the challenges related to ensuring rigor and comparability for measuring student growth in non-tested grades and subjects;
- Help states develop consistent, reliable, and appropriate teacher observation instruments for performance-based teacher evaluations; and
- Consider potential solutions and opportunities for state collaboration to address the challenges states face in designing and implementing comprehensive teacher and principal effectiveness systems.

This paper synthesizes some of the expert presentations and discussions among Race to the Top state grantees during six events listed below which occurred during the first six months of the Teacher and Leader Effectiveness CoP:

- November 10, 2010 webinar entitled “Getting the Math Right: Aligning Value-Added and Student Growth Models to State Policy Expectations,” which featured Dan Goldhaber, Ph.D.,

director of the Center for Education Data and Research (CEDR) and Daniel McCaffrey, Ph.D., a senior statistician at the RAND Corporation.

- November 17, 2010 webinar entitled “Non-Tested Grades and Subjects: Options for Measuring Student Growth,” featuring Robert Meyer, director of the Value-Added Research Center and professor at the Wisconsin Center for Education Research and William Slotnik, founder and executive director of the Community Training and Assistance Center.
- December 10-11, 2010 in-person convening on teacher and leader effectiveness, featuring presentations on teacher observation instruments and tools by a number of technical experts (see Appendix A), including Courtney Bell, Ph.D., Research Scientist at Educational Testing Service (ETS); Tim Daly, President of The New Teacher Project; and Dr. Tony Bryk, President of the Carnegie Foundation for the Advancement of Teaching.
- April 14, 2011 state peer webinar entitled “Measuring Student Learning in Educator Evaluation: Rhode Island’s Model Under Development,” led by state officials from Rhode Island.
- May 5, 2011 state peer webinar entitled “Value-Added and Student Growth Models: Operating Rules,” led by state officials from Tennessee.
- May 18-19, 2011 in-person meeting of the Reform Support Network’s Teacher and Leader Effectiveness Community of Practice, focused specifically on options and considerations related to measuring student growth in non-tested grades and subjects.

The considerations below aim to share broadly across *all* states some guiding questions and expert thinking on the early challenges and critical decisions facing states building performance-based teacher evaluations as part of systems of educator effectiveness. The topics include:

- 1) **Placing teacher evaluation design** in the context of state goals for framing a comprehensive educator effectiveness system;
- 2) **Choosing value-added and/or student growth models** to measure teacher impact on student achievement;
- 3) Developing **student growth measures in subjects and grades not covered by required statewide assessments**; and
- 4) **Choosing appropriate observation instruments** for teacher evaluations.



## **Considerations on Building Performance-Based Teacher Evaluations**

### **Looking at the Big Picture: Framing the Design of Teacher and Leader Effectiveness Systems**

Developing a system of educator effectiveness is a complex undertaking. There are many pieces that must fit together, from data systems for measuring student performance and linking teacher and student data, to teacher and principal evaluation, teacher and principal preparation, compensation, and professional development design.

According to Dr. Tony Bryk, President of the Carnegie Foundation for the Advancement of Teaching, and an expert on systems change, the first key step for a state is to develop clarity about goals, purposes, cost, and capacity, and to implement a comprehensive process that allows the state to engage stakeholders, develop a theory of action, plan and prepare for next steps.

Before focusing on any individual work streams or activities, it is beneficial for states to invest significant time and effort to identify and prioritize reforms that are most likely to improve the effectiveness of their teachers. This requires states to first consider the “big picture” view of the work and assess the ways that various state and district policies affecting teacher quality work together (or do not work together) to achieve a comprehensive educator effectiveness system.

Dr. Bryk observes that the rules and regulations now being developed by states and districts incorporate many technical details regarding the measurement of teacher performance through evaluations. These new rules and regulations will, in turn, shape practices in classrooms, schools, and district offices in profound and potentially unexpected ways. Success in improving student learning will depend on how effectively the field is able to understand and articulate the assumptions behind these rules and regulations, and integrate technical, regulatory, and practical considerations into a system that is continuously improving.

Bryk recommends that states consider the following questions as they begin a process of rethinking teacher evaluations in the context of a coherent and comprehensive educator effectiveness system:

- How do teacher evaluations **fit into the larger educator effectiveness system** or set of policies in place to ensure that the state recruits, develops, supports and retains highly-effective teachers?
- What are **the relationships, synergies, and dependencies** among teacher standards, teacher measures, teacher training, teacher qualifications, and policies that govern tenure, dismissal, promotion, compensation, and professional development? How coherent and aligned are these policies?

- What are the **desired outcomes** of the educator effectiveness system and what are the key policy levers that will lead to those outcomes?
- Given that the system is intended to do more than identify the lowest performing teachers, what are the **system strategies for improving teacher performance** more broadly?
- What do teachers, principals, and other leaders **believe about the ultimate purpose** of the teacher evaluations in particular and the state's educator effectiveness system in general?

Experts note that it is critically important for states to provide time and opportunities for stakeholders to consider these important framing questions. With input from stakeholders on these issues, each state will be better prepared to articulate a theory of action that clearly describes its approach to developing performance-based teacher evaluations as part of a comprehensive system of educator effectiveness.

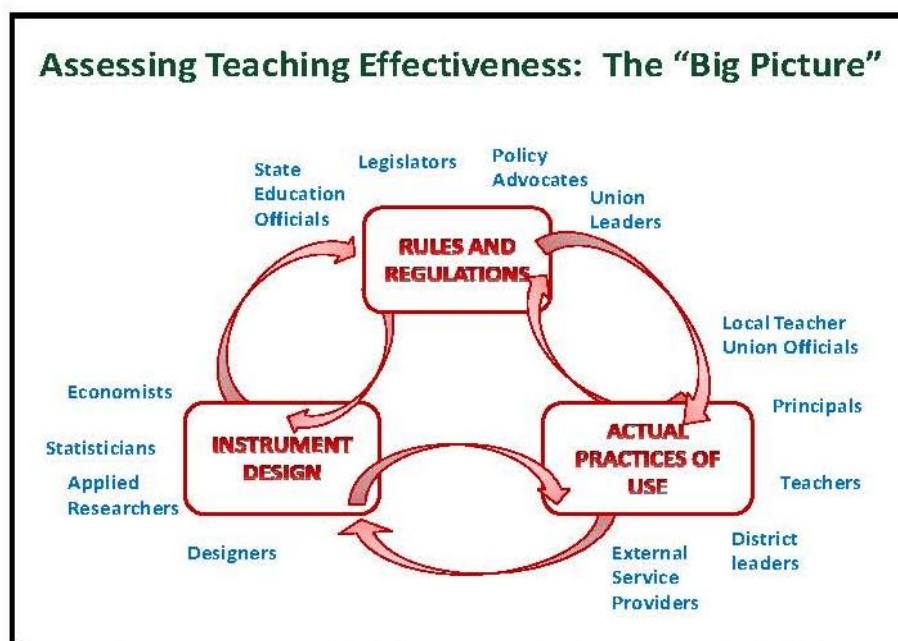
---

### Framing Educator Effectiveness System Design

- *Maintain a “big picture” view of the work.*
- *Be clear how efforts to implement effectiveness systems and the need for better teacher evaluations are part of a larger effort to improve the professional and daily working life of teachers.*
- *Articulate the “theory of action” behind the state’s approach to teacher evaluation.*
- *Take an “improvement research orientation” to the work; that is, approach building teacher and leader effectiveness systems with flexibility and willingness to adjust policy based on experience, data, and feedback from research on reforms in action.*

*Involve stakeholders – including teachers, state level education officials, legislators, union leaders, district and school leaders, experts and researchers – early and consistently.*

---



**Source:** From a presentation titled, “Assessing Teaching: Seeing the Full Context,” by Dr. Tony Bryk at the December 10-11, 2010 in-person convening on teacher and leader effectiveness.

**Developing a theory of action.** A state’s answers to the framing questions above provide the framework for articulating a theory of action for how the state will design its teacher evaluations within the context of a broader educator effectiveness system. The theory of action unpacks the decisions for various aspects of the evaluation design – from purposes and goals to design and outcomes. Each of the decision points outlined along the way in the design and the development of teacher evaluations must align with a state’s theory of action.

**Framing teacher effectiveness.** Experts broadly agree that policymakers and system developers also need to consider how ways they frame discussions of effectiveness will be received by teachers and school leaders and how new policies could affect their daily work. Bryk asserts that state and district policymakers need to recognize that although challenges in the classroom and district are important, it is the school factors – the characteristics of the work environment – that have the most significant impact on whether a teacher remains in the profession. This is a critical consideration for educator effectiveness system design, given that many aspects of implementation occur in the school and affect teachers’ work environments.

**Adopting an improvement research orientation.** Bryk also argues that it is important for states to take an “improvement research orientation” to the work of building teacher evaluations and educator effectiveness systems. States must also convey to all stakeholders, from the beginning, that development must be a process of continual improvement. It is important for stakeholders to acknowledge at the outset the importance of continuously reflecting on the progress of the work and making adjustments based on what is learned. Building mechanisms to evaluate progress from the beginning and setting a standard of continual improvement for the state’s efforts may be crucial for success.

**Assessing costs and capacity.** Finally, Bryk recommends that states inventory and carefully consider the costs, resources, and capacity needed to make a performance-based teacher evaluation system and, more broadly, an educator effectiveness system, function well over time. This will require states, districts, and schools to identify new roles and responsibilities and plan for the intensive training and support that will be needed to carry out those new roles and responsibilities.

## Focusing on Student Achievement: Choosing the Right Student Growth Model

Efforts to measure student growth and use data to inform education decision-making are not new. A few states have already implemented models to measure student growth for use in state accountability systems and adequate yearly progress determinations. But in designing and implementing performance-based teacher evaluations, many states now plan to go further, using measures of student growth as (a significant) one of multiple measures to inform their assessments of teachers and leaders.

With a big picture theory of action in place, there are numerous technical challenges that states face in developing teacher evaluations that are grounded in student performance. One of the key decision points that states face in developing performance-based teacher evaluations is choosing the method by which student achievement data are linked to individual teachers to make inferences about teacher performance and effectiveness.

In *Measuring Teacher Effectiveness Using Growth Models: A Primer*, the Reform Support Network discusses the important differences between value-added models and student growth percentile models.<sup>4</sup> States must consider these differences carefully.

**Value-added models (VAMs)** are a specific type of growth model in the sense that they are based on changes in test scores over time. However, not all growth models are VAMs. VAMs specifically attempt to determine how specific teachers and schools affect growth in student achievement over time. VAMs are relativistic, addressing the question – to what extent can changes in student performance be *attributed to a specific school and/or teacher compared with that of the average school or teacher?* VAMs are complex statistical models that generally attempt to take into account student or school background characteristics in order to isolate the amount of learning attributable to a specific teacher or school. Teachers or schools that produce more than typical or expected growth are said to “add value.”

---

<sup>4</sup> The text in this section is drawn from *Measuring Teacher Effectiveness Using Growth Models: A Primer*, which was prepared by the Reform Support Network and is available at: [http://www.swcompcenter.org/educator\\_effectiveness2/Measuring\\_Teacher\\_Effectiveness\\_Using\\_Growth\\_Models.pdf](http://www.swcompcenter.org/educator_effectiveness2/Measuring_Teacher_Effectiveness_Using_Growth_Models.pdf)

There are a variety of VAMs, but they can be categorized into the following major groups:

- **Gain score models:** measure year-to-year change by simply subtracting the prior year score from the current year score; typically the gains for all students for a given teacher are averaged.
- **Covariate adjustment models:** model current year test scores as a function of the prior year test scores and other student and classroom characteristics.
- **Layered models (including the persistence model):** model scores for multiple years in multiple subjects that may or may not include student background variables.

**Student growth percentile models** are another type of growth model that may be used to examine the contribution of teachers to student growth. In this model, a different type of statistical procedure is used to examine *changes in student achievement for individual students compared with other students*. This information is then aggregated to the teacher level to produce an estimate of the teacher's impact on student learning.

Each of these approaches has different characteristics, advantages, and challenges. To date, there is no consensus among experts on the “best” model. Value-added analysis involves a series of steps from collecting and organizing data to deciding on a model and implementing and reporting on the results of that model. Each step is important and can have significant consequences for individual teachers and schools.

What experts do agree on is that selecting a model—whether VAM or student growth percentile model—involves states thinking carefully about what types of decisions will be made with the results and what model will provide the best information for these decisions. Not all models will necessarily produce equally useful information for each type of decision.

Ultimately, policymakers must ask themselves how credible and useful their model will be to their stakeholders and toward their student achievement goals. Issues states could think about in examining these models include the following:

- **Precision.** How precise is the model? How does the model account for error? How well will it differentiate among teachers? How likely is the model to misidentify teachers as highly-effective or ineffective (i.e., false positives or negatives)?
- **Validity.** What evidence is there that results from the model align with other measures of teacher effectiveness?
- **Data needs.** How much data and what type of data are needed to implement the model? Does the state have these data readily available? How does the model handle missing data? How does the model deal with student mobility?
- **Fairness and expectations for student achievement.** What factors or variables does the model include as controls? Why? How do the costs of including these controls (e.g.,

increasing model complexity, possible implications of differing expectations for different students, increasing possibility of more missing data) compare with the benefits (e.g., accounting for differences in student characteristics that are not attributable to teachers and schools, possibly improving precision)?

- **Stability and changes in estimates.** How many years of data are included for each estimate? Does the model estimate year-to-year change, or does it average information over multiple years? What are the tradeoffs in terms of precision and stability compared with the potential to see changes in estimates over time?
- **Comprehensibility of model and results.** How easy is the model to explain and describe to stakeholders, such as teachers? How are results typically presented, and how will they compare to measures of status or raw growth?
- **Cost.** How much will it cost to implement the model over time? Specifically, what are the upfront costs of development and the ongoing costs of system maintenance and improvement?
- **Ease of implementation and ownership.** What capacity (psychometric, software, etc.) is needed to implement the model? Can the state implement or “own” the model over time?
- **Alignment with other measures.** How does the model align with existing growth or value-added models in place in the state?
- **Usability.** How easily can data from the model be used along with other data to assess teacher or school leader practice?
- **Ability to aggregate to school level.** How can information from this model be used to help evaluate principals or other school leaders? Can the model or a different specification of the model estimate principal or school effects? How? What is the interpretation of these estimates?

With a strong understanding of the strengths, weaknesses, and trade-offs among various student growth models, states can then return to their basic theories of action to make choices about how to measure this key component of performance-based teacher evaluation. In addition, states need to begin to define the basic operating rules of the systems they are designing. To do this, states could:

**Consider the policy outcomes.** What decisions are the value-added or student growth models intended to inform? Different policy decisions, such as decisions about professional development, rewards, dismissal, tenure, or school accountability, require different data. States can backward map from their goals and their theory of action to help ensure that the purposes for which the state wishes to use data shape the specifications.

**Engage stakeholders.** Work with peers and technical experts to design a plan for engaging stakeholders in the selection of the model. Be transparent by documenting every step toward



decisions related to selection of a model so that decisions can be tracked, explained, and communicated publicly.

**Ensure consistency and capacity.** For assessment data that are comparable statewide, it is important that state and district data collections are consistent and that there is a capacity to generate the information needed for linking students and teachers. State trend data can also be used to inform decisions around cut-off points, and data from prior years can be useful in assessing the system before implementation. In addition, longitudinal data from other states can be used to inform decisions. States can also work with districts to gain insight into the system's potential strengths and limitations, especially in cases where the data will be used to inform decisions about individual teacher performance.

---

### **A Tutorial on Assessing Student Performance and Growth By the Value-Added Research Center**

*For a useful and user-friendly tutorial on the similarities and differences among models for assessing student performance—attainment, growth, and value added—see Value-Added Research Center at <http://varc.wceruw.org/tutorials/>*

*The tutorial uses the analogy of gardeners growing trees to illustrate the different ways of looking at student achievement and how that information might be used to assess teacher performance. The tutorial could be a useful tool for engaging stakeholders in forums regarding how growth models can help inform student achievement and help build a system of educator effectiveness.*

---

**Take critical steps to ensure sustainability.** It is important for states to plan and prepare for continual monitoring and refinement. Understanding state and local capacity is essential. Who will assume responsibility for managing different aspects of the measurement system? Who will be responsible for data quality and by what process will data quality be ensured?

**Define the operating rules of the teacher evaluation system.** What are the rules for using the chosen models, considering state and local data system capacity, and identifying the ways in which the models will contribute to measuring teacher effectiveness? For example, which students and what achievement data should count toward a teacher's performance evaluation? How much will growth measures count towards a teacher's overall evaluation? Will growth data for individual teachers be made public? How will student scores be attributed to teachers and schools? How will growth be measured for students who are highly mobile? How should the model account for student background characteristics?

### Example: Some of Tennessee's Operating Rules

*Tennessee's Value-Added Assessment System (TVAAS) has been in place since 1992. The state has statewide standardized Tennessee Comprehensive Assessment Program (TCAP) assessments in reading, math, science, and social studies in grades 3-8. In 2010, Tennessee implemented a new data dashboard and began training around TVAAS teacher effect scores with a plan for new annual teacher and principal evaluations for 2011-12. Operating rules include:*

- *Teacher effect estimates are mandated by state statute;*
- *Annually, data from the TCAP tests, or their future replacements, will be used to provide an estimate of the statistical distribution of teacher effects on the educational progress of students within school districts for grades (3-8);*
- *A student must have been present for 150 days of classroom instruction per year or 75 days of classroom instruction per semester before that student's record is attributable to a specific teacher;*
- *The estimates of specific teacher effects on the educational progress of students will not be public record, and will be made available only to the specific teacher, the teacher's appropriate administrators as designated by the local board of education, and school board members; and*
- *Thirty-five percent of the evaluation criteria shall be student achievement data based on student growth data as represented by the TVAAS or some other comparable measure of student growth, if no such TVAAS data are available.*

**For more information on Tennessee's teacher evaluation plans, see**  
<http://www.tn.gov/firsttothetop/>.

---

**Prepare a communication strategy.** An effective communication strategy around incorporating student achievement into performance-based teacher evaluations will explain to all stakeholders the intent, goals, strengths, and limitations of value-added and student growth models. Most often, these models evaluate teachers on their contribution to student growth rather than overall proficiency, and recognize that there are factors contributing to student proficiency which are beyond a teacher's control. Many experts argue that these models may be much fairer to teachers because they usually take a student's background and prior performance into consideration – a fact often missing in recent debates on whether it is fair to judge teachers based on the performance of their students. States might consider conducting strategic planning with small groups of stakeholders to establish common agreements that can pave the way for constructive engagement with all stakeholders as work progresses.



## Rigor and Comparability: Developing Growth Measures in Non-Tested Grades and Subjects

Value-added and other growth models are most readily suited to situations where standardized student assessment data are available. Statewide tested grades and subjects afford relatively large and robust data sets that can be used to measure changes in student academic achievement.

In most states, because of Elementary and Secondary Education Act (ESEA) assessment requirements, statewide data are readily available for many students and teachers in grades 3-8, as well as high school math and English/language arts.

However, even with these requirements, statewide standardized assessment data may not be available for more than half of the teachers in a given state. Thinking about the full complement of teachers – *including K-2, social studies, special education, non-core subject areas, and teachers of English Language Learners* – states face the challenge of how to develop fair, rigorous, and comparable measures of student growth and achievement that can be used to evaluate the performance of teachers for whom state standardized achievement data do not exist.

Given this, how should a state approach developing student growth measures in grades and subjects for which there are no statewide standardized assessments? When measuring student growth in “non-tested grades and subjects” (NTGS), other measures need to be used or developed.

The Reform Support Network’s *Measuring Student Growth in Non-Tested Grades and Subjects: A Primer* identifies three general approaches emerging from state and district practice as well as expert thinking in response to the challenge of measuring student learning in NTGS.<sup>5</sup> It is important to note that these approaches are not mutually exclusive. It is likely that states and districts may want to use a variety of approaches to measuring student growth depending on the assessments available, the costs and benefits of each approach, and the contextual needs within the state. Examples of these approaches include:

- **Student learning objectives (SLOs)** are a participatory method of setting measurable goals, or objectives, based on the specific assignment or class, such as the students taught, the subject matter taught, the baseline performance of the students, and the measurable gain in student performance during the course of instruction. SLOs can be based on standardized assessments, but they also may be based on teacher-developed assessments or other classroom assessments if they are “rigorous and comparable across classrooms.” When using SLOs, teachers set measurable expectations for student learning, usually in collaboration with their principal or other leader. SLOs can also be used in tested grades and subjects to help determine how predictive the measures of student growth are, and using them in all grades and subjects assures some comparability in methods.

---

<sup>5</sup> The text in this section is drawn from *Measuring Student Growth in Non-Tested Grades and Subjects: A Primer*, which was prepared for the Reform Support Network and is available at [http://www.swcompcenter.org/educator\\_effectiveness2/NTS\\_\\_PRIMER\\_FINAL.pdf](http://www.swcompcenter.org/educator_effectiveness2/NTS__PRIMER_FINAL.pdf).

### Student Learning Objectives: Rhode Island

In order to ensure that objectives are specific, measurable, and rigorous, evaluators must establish clear processes for setting them. Rhode Island, an example of a state using SLOs for NTGS, recommends that states consider how:

- District leadership and building administrators take time to establish a process of setting SLOs that ensures objectives are aligned to district and school goals.
- Processes should be established such that, at a minimum, teachers in a school who teach the same grade/subject have the same objectives and evidence (but may have different “targets” depending on their “baselines”).
- Eventually, teachers in different schools who teach the same content have similar objectives and comparable evidence.

Rhode Island also notes the importance of checks and balances when using SLOs to ensure that they are rigorous and evaluated objectively. This includes regular audits of principal evaluations of teachers and state training of a cadre of intermediary service providers (often experienced retired teachers and principals) to undertake evaluations.

---

- **New or existing measures of student growth** (including pre- and post-tests as well as performance and portfolio assessments) can be used to measure student growth in non-tested grades and subjects. These measures may include early reading assessments; end-of-course assessments; and benchmark, interim, or unit assessments. Other assessments may be developed at either the state or district level.

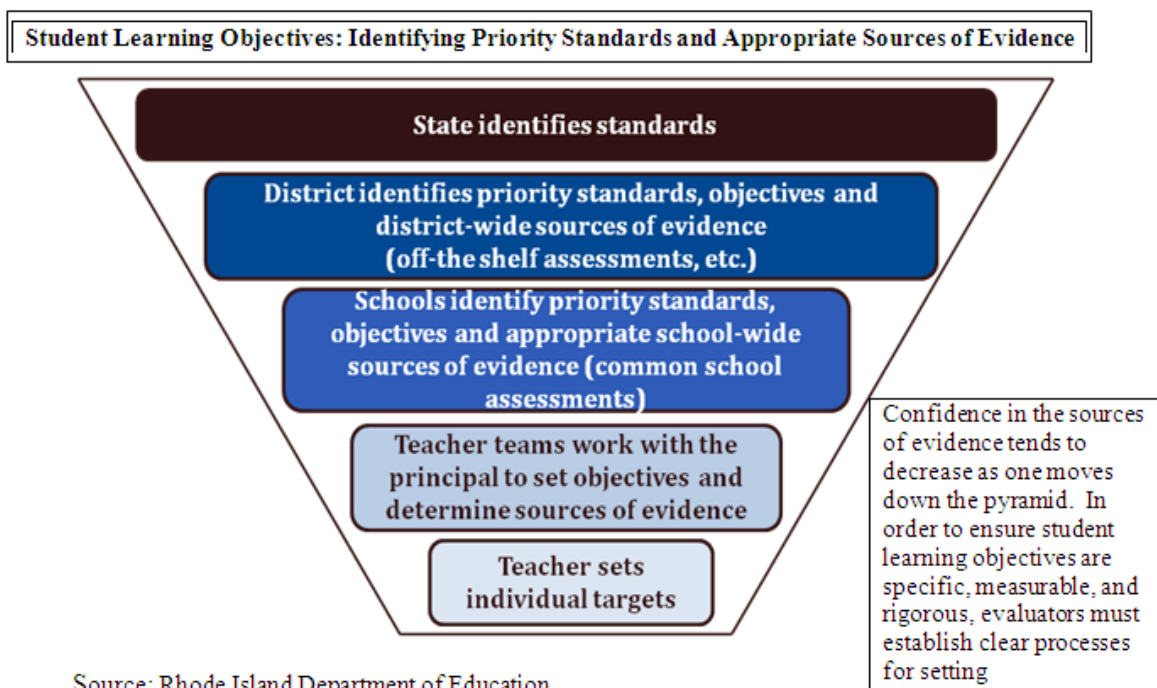
State discussions within the Teacher and Leader Community of Practice (CoP) make it clear that for some states, new options for measuring growth, such as those listed above, are critical. A number of states are interested in working towards new approaches to measuring student growth for grades K-2, social studies, special education and for English language learners. Options that states are exploring include development of early reading assessments aligned with the Common Core State Standards or progress monitoring protocols in the early grades. Some states see opportunities for collaboration across states in assessing middle school subjects such as social studies.

In each case, a goal for new assessment options is to increase the amount of comparable student learning data available for use in a broader system of educator effectiveness that differentiates and tailors professional development and improves student outcomes.

- **Measures of collective performance** assess the performance of the school, grade, instructional department, teams, or other groups of teachers. These measures can take a variety of forms including school-wide student growth measures, team-based collaborative achievement projects, and shared value-added scores for co-teaching situations.

As states consider selecting an approach to measuring teacher effectiveness in grades and subjects for which there are no statewide standardized assessment data, they are engaged in a process of weighing the costs and benefits of different options.

Similar to the process for considering value-added and growth models, states are examining the desired policy outcomes of the models selected; working with peers and technical experts to devise plans to engage stakeholders in the selection of the approaches for non-tested grades and subjects; defining operating rules; considering state and local data system capacities; identifying the ways in which the models will contribute to measuring teacher effectiveness; ensuring that state and district data collections are consistent; planning and preparing for continual monitoring and refinement; and preparing communication strategies that explain to educators and the public the intent, goals, strengths, and limitations of student growth modeling in non-tested subjects and grades.



In devising growth measures to assess teacher performance in subjects and grades for which no state standardized data are available, experts recommend that states consider:

- **Using existing assessment tools that are already available and appropriate** for this purpose. The Center for Educator Compensation Research is “in the process of developing a census of assessments being provided by all US states and selected districts. [Its] goal is to capture innovative approaches these education agencies have taken to implement assessments

in grades, subjects, and languages, not required under [ESEA].”<sup>6</sup> This resource may provide useful information on available assessments that could be adopted or adapted for state or district use in current non-tested grades and subjects. Of course, states would have to carefully examine whether proposed assessments are aligned with state standards and serve their own goals.

- **Working strategically with vendors** to select or create a test bank of assessment items for current non-tested grades and subjects that are aligned with state standards and are appropriate for the purposes for which they will be used.
- **Identifying opportunities for collaboration** with other states, such as sharing item banks, determining best practices, and creating common assessments.
- **Engaging teachers in developing new assessments** for measuring student growth in non-tested grades and subjects. Many subjects have standards to guide assessment development, and national organizations often offer assessments for specific subjects. Teachers can bring a wealth of knowledge and ready examples to these discussions.
- **Incorporating predictors into the value-added or student growth model** for all grades and subjects. States may consider using secondary predictors such as college entrance exam scores or other tests, in addition to pre-tests, for a specific subject, to determine predicted growth.
- **Providing support for districts** on developing consistent measures or rigorous student learning objectives. States might consider developing models and prototypes of learning standards, sharing exemplars on their websites, as well as providing written guidance and decision-making tools to districts on the state’s standards for building student growth expectations for subjects and grades for which no statewide data are available.
- **Maintaining quality control** at the state level for how growth is measured in NTGS. Some states are requiring districts to submit their plans and methodology for developing growth measures for NTGS. By having a vetting process or state audit in place, states can help ensure that districts make good faith efforts to measure teacher performance in NTGS in a fair and reliable manner. States have some tools at their disposal for considering whether locally developed or proposed NTGS measures are defensible. The first is to use standardized statewide measures as a basis for comparison. To what degree do judgments about teacher performance in NTGS resemble the pattern in teacher performance on standardized statewide measures? Second, does the NTGS measure result in differentiation – does it result in a distribution of teacher performance which, at the very least, distinguishes between the best and worst teacher performance across the spectrum of teachers evaluated?

---

<sup>6</sup> See <http://www.assessmentsurvey.wceruw.org/>.

**Teacher and Leader Community of Practice:  
Considerations for Student Growth Measures in Non-Tested Grades and Subjects**

- *Are existing assessments consistent and comparable? Do they allow for measurement of student progress over time?*
  - *Should existing tests be used? Will additional assessments or new measures of student learning need to be developed?*
  - *What process will states use to ensure that locally developed measures of student growth are credible and reliable measures of teacher effectiveness?*
  - *What strategies need to be developed to ensure that there is meaningful engagement with stakeholders?*
  - *What are the costs associated with the various approaches?*
  - *What are the data capacity requirements for measuring growth in non-tested grades and subjects?*
- 

Regardless of the strategies states pursue to measure growth in NTGS, experts emphasize that it is important for states to *prioritize*. It is important for states to develop fair and reliable student growth measures for which statewide assessment data *are* available. After that, states and districts can and should prioritize based on enrollment counts, number of teachers by subject and grade level, as well as the availability of consistent student achievement measures, to address the student achievement portion of teacher evaluations in non-tested grades and subjects.

Moving towards comparable measures of student growth to use in teacher evaluations for NTGS is important, and getting comparability within a district for teachers within the same grade and subject area is itself a substantial accomplishment – but it is not the only goal. Experts note the importance of ensuring that measures used to evaluate teachers are *rigorous* and *fair* for every teacher.

One consistent theme in state discussions on this issue is how important *transparency* is in developing measures of teacher performance – a point that is true for all aspects of teacher evaluations. In particular, where there may be critical questions about comparability, it is essential that states and districts can make clear cut cases for the fairness, the rigor, and the appropriateness of the measures chosen for evaluation.

### **Example: Delaware's Teacher Evaluation System**

Delaware has had a statewide educator evaluation system since the 1980s. The state's current evaluation system, the Delaware Performance Appraisal System (DPAS) II, has been in use since 2008. It includes three versions, one for administrators, one for teachers, and one for specialists.

DPAS II for teachers and specialists is based on Charlotte Danielson's *Enhancing Professional Practice: A Framework for Teaching (2nd Edition)*, while DPAS II for administrators is based on the Interstate School Leaders Licensure Consortium's (ISLLC) standards for leaders.

For all educators, DPAS II defines standards for professional practice along five components: 1) planning and preparation, 2) classroom environment, 3) instruction, 4) professional responsibilities, and 5) student improvement. For each of the first four components, there is a set of four appraisal criteria. Each criterion has a rubric defining "unsatisfactory," "basic," "proficient," and "distinguished" performance.

Evidence for performance on components 1, 2, and 3 for teachers and specialists is gathered through observations by administrators trained in assessment. Evidence for performance on components 1, 2, and 3 for administrators is gathered through a survey completed by professional staff, the administrator's self-assessment on the ISLLC standards, and the assessor's survey data. For the fourth component, all educators complete a professional responsibilities form, which details their professional growth, communication with students, parents, and school colleagues, and their contributions to the professional community during the review period.

To receive a "satisfactory" rating for each of the first four components, an educator must receive a satisfactory ("basic," "proficient," or "distinguished") on at least three of the four criteria specified in the component.

Under Delaware's recently revised regulations, beginning in July 2011, a satisfactory rating for the fifth component (student improvement) means that the teacher has met the standard for student growth. That standard will represent an appropriate level of change in achievement data for an individual student between two points in time, as well as any other measures that are determined to be rigorous and comparable across classrooms.

Currently, assessments can result in summative ratings of "effective," "needs improvement," or "ineffective." Under the revised regulations, Delaware will add a fourth summative rating of "highly effective" in July 2011. Educators will be required to demonstrate satisfactory levels of student growth to receive an "effective" rating, and more than a year of student growth to receive a "highly effective" rating.

**For more information on Delaware's Race to the Top plan, see**  
<http://www.doe.k12.de.us/rttt/>

---

## Observing Teacher Practice: Choosing Appropriate Classroom Observation Instruments

Teacher evaluations have historically used some kind of classroom observation to obtain information about teacher effectiveness. Observations can provide important information that can be used to support professional growth, improve performance, and make decisions about compensation, employment, and other aspects of an educator effectiveness system. As states are assessing their current tools or considering new options, they should align the purposes and methods of observation with the expectations of their educator effectiveness system, and build in methods to assess inter-rater reliability among the individuals tasked with conducting teacher observations.

In choosing observation instruments to incorporate into performance-based teacher evaluations, experts in the field urge states to:

- **Clarify the purpose and objectives** of the teacher evaluation system and how observation instruments can help meet those objectives;
- **Evaluate the rigor, quality, and utility** of observation instruments under consideration;
- **Engage stakeholders**, including teachers and principals, in the design and selection of observation instruments; and
- **Provide professional development** to principals, teachers, and other raters to ensure that observation instruments are implemented with fidelity.

---

### Teacher and Leader Community of Practice: Challenges Related to Observation Instruments

- *Ensuring that the observation instruments are aligned with the purposes and methods of the teacher evaluation system;*
  - *Ensuring that teacher evaluations are conducted in a consistent and reliable manner by evaluators;*
  - *Helping evaluators provide high-quality feedback to teachers that will help them improve their teaching; and*
  - *Establishing a common language on instructional practice to help district leaders develop consistent and effective professional development for teachers.*
-



Courtney Bell, Research Scientist for Educational Testing Service, recommends that states consider the research-based observation instruments already in use, including, for example:

Instrument	Developer(s)	Subject Area(s)	Grades
Framework for Teaching (FFT)	Charlotte Danielson	All	K-12
Classroom Assessment Scoring System - Secondary (CLASS-S)	Bob Pianta & Bridget Hamre	All	6-12
Mathematical Quality of Teaching (MQI)	Heather Hill, et al.	Math	4-12
Protocol for Language Arts Teaching Observations (PLATO)	Pam Grossman, et al.	English/Language Arts	4-12
Quality of Science Teaching (QST)	Ray Pecheone, et al.	Science	6-12

The table above includes some examples of existing instruments but is certainly not exhaustive. Bell provides several important indicators for assessing the quality of teacher observation protocols, applicable to any teacher observation instrument:

- There is a **clear articulation of score use**.
- There are **meaningful and observable differences** between score points.
- The **inferences required of the rater can be made reliably**, given training and support.
- There is **validity evidence** to support use of the instrument.
- **Teachers understand the scales** and score point distinctions.
- **Raters score consistently and accurately** at acceptable levels (~80%).
- Observational score and other quality indicator **comparisons make sense**.

Tim Daly, President of The New Teacher Project, also suggests that states consider existing instruments for observing teachers as part of performance-based teacher evaluations, but cautions against individual states or districts making modifications that alter the psychometric properties and validity of the instruments. He also suggests that in order for states to determine whether the observation instrument, its criteria, and its tools would contribute to accurate evaluation results, four key questions should be answered:

- **Is the instrument grounded in what matters to student achievement?** Does the instrument consider the classroom performance areas most connected to student outcomes, such as lesson objectives; strategies, activities, and delivery; physical environment;



classroom management and leadership; student engagement and real-time assessment; end of class assessment; and student mastery of lesson objectives?

- **What expectations does the instrument set?** Does the instrument set high performance expectations for teachers or outline only minimally acceptable performance?
- **Are the performance expectations for teachers unambiguous and precise?** That is, are the performance expectations clear enough that they leave little room for interpretation, telling observers exactly what to look for, or are they vague and general, leaving too much room for interpretation?
- **Is the instrument student-centered?** Does it require evaluators to look for direct evidence of student engagement and learning? Some observation tools focus only on the teacher's skills and behavior, without also including a focus on student response and impact, as well.

Experts also note that while implementing observation systems with fidelity requires significant time and expense, technology has the potential to ease costs and other challenges. Video databases, for example, are a potentially important emerging technology for evaluator and teacher training, evaluator (re)calibration, professional development, and principal workload management.

## **Conclusions and Looking Forward**

Teacher evaluation is just one of several critical areas to consider in developing a coherent and aligned system for educator effectiveness –this paper only begins to explore some of the key issues and challenges states face in designing teacher evaluations that can identify varying levels of instruction and provide actionable information on improving teacher practice and ultimately, student achievement. Both state experience and expert advice on implementing performance-based teacher evaluations and systems of educator effectiveness suggest the following emerging guidelines:

- Ultimately, performance-based teacher evaluations are meant to be part of an overall educator effectiveness system **dedicated to improving student learning outcomes**.
- Educator effectiveness systems should be **built with the intention of improving individual and collective practice**, and facilitating the overall growth of the workforce of teachers and leaders. The information developed by and used in these systems should identify both strengths and weaknesses of individual teachers; provide rich information about students; be provided in a timely, user-friendly format to teachers and school leaders; and be used as the basis for policy and personnel decisions designed to improve student and school performance.
- As states move forward they are addressing **how teacher performance measures and instruments** align with teacher policy in other important aspects of a comprehensive system of educator effectiveness, including areas such as preparation, recruitment, tenure, promotion

policy, professional development, and policies focused on ensuring an equitable distribution of effective teachers across schools.

With so many states engaged in this complex and technical work simultaneously, the Reform Support Network has a special focus on helping Race to the Top states identify opportunities to collaborate with each other and on sharing those lessons broadly. In a recent meeting of the Teacher and Leader Effectiveness Community of Practice, a variety of Race to the Top states identified the following areas as priorities for potential state collaboration and solution design:

- Guidance on assessment design, especially for district, school, or classroom measures that might be used for evaluation in non-tested grades and subjects;
- Development of valid and reliable measures of student growth for kindergarten through second grade;
- Development of a cross-state item bank (particularly for non-tested grades and subjects); and
- Building a generic state framework to guide processes for developing and implementing student learning objectives.

In addition, the Reform Support Network is exploring ways that states can overcome resource barriers to reforms through cost-sharing and group purchasing strategies. Convening Race to the Top states to discuss solutions to common technical challenges is just one approach of the Teacher and Leader Community of Practice. The Network will continue to pursue opportunities to assist grantees in improving educator effectiveness and student achievement by providing resources that will help address challenges collaboratively and effectively. As the work progresses, the Reform Support Network will continue to listen actively to state needs and will regularly share key learnings and best practices with all states.

## Appendices

### Appendix A – List of Technical Experts Participating in the Teacher and Leader Effectiveness Community of Practice

Courtney Bell, Educational Testing Service  
Tony Bryk, Carnegie Foundation for the Advancement of Teaching  
Steve Cantrell, Bill & Melinda Gates Foundation  
Tim Daly, The New Teacher Project  
Ben Fenton, New Leaders for New Schools  
Laura Goe, Educational Testing Service  
Dan Goldhaber, Center for Education Data and Research, University of Washington  
Brian Gong, National Center for the Improvement of Assessment and Harvard University  
John Hussey, Battelle for Kids  
Thomas Kane, Bill & Melinda Gates Foundation  
Richard Laine, The Wallace Foundation  
David Lussier, Austin Independent School District  
Dan McCaffrey, RAND Corporation  
Robert Meyer, Value-Added Research Center, University of Wisconsin  
Richard Pennington, Scope Vision  
Bill Slotnik, Community Training and Assistance Center (CTAC)  
Chris Thorn, Value-Added Research Center, University of Wisconsin

### Appendix B – Relevant Resources

Community Training and Assistance Center. 2008. *CMS Student Learning Objective Guide*. Boston, MA: Author. Retrieved January 20, 2011, from [http://gse.berkeley.edu/research/pace/reports/altcomp/Smith\\_Student\\_Learning\\_Guide.pdf](http://gse.berkeley.edu/research/pace/reports/altcomp/Smith_Student_Learning_Guide.pdf).

Community Training and Assistance Center. 2004. *Catalyst for Change: Pay for Performance in Denver Final Report*. Boston, MA: Author. Retrieved January 20, 2011, from <http://www.ctacusa.com/PDFs/Rpt-CatalystChangeFull-2004.pdf>.

Goe, L., C. Bell, and O. Little. 2008. *Approaches to Evaluating Teacher Effectiveness: A Research Synthesis*. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved January 20, 2011, from <http://www.tqsource.org/publications/EvaluatingTeachEffectiveness.pdf>.

Holdheide, L. R., L. Goe, A. Croft, and D. J. Reschly. July 2010. "Challenges in evaluating special education teachers and English language learner specialists." *TQ Research & Policy Brief*. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved January 20, 2011, from <http://www.tqsource.org/publications/July2010Brief.pdf>.

- Little, O., L. Goe, and C. Bell. 2009. *A Practical Guide to Evaluating Teacher Effectiveness*. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved January 20, 2011, from <http://www.tqsource.org/publications/practicalGuide.pdf>.
- Milanowski, A. T., G. G. Henemann, and S. M. Kimball. 2009. *Review of Teaching Performance Assessment for Use in Human Capital Management*. Madison, WI: Strategic Management of Human Capital-Consortium for Policy Research in Education. Retrieved January 20, 2011, from <http://www.smhc-cpre.org/download/69/>.
- National Comprehensive Center for Teacher Quality. January 20, 2011. *Guide to Teacher Evaluation Products*. Retrieved from <http://www3.learningpt.org/tqsource/GEP/>.
- National Council on Teacher Quality. 2010. *Blueprint for Change: National Summary* at [www.nctq.org/stpy](http://www.nctq.org/stpy).
- The New Teacher Project. 2010. Evaluation 2.0 at <http://tntp.org/publications/issue-analysis/view/teacher-evaluation-2.0/>.
- The New Teacher Project. 2009. *The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness*. See <http://widgeteffect.org/>.
- Steiner, L. October 2009. "Determining Processes That Build Sustainable Teacher Accountability Systems." *TQ Research & Policy Brief*. Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved January 20, 2011, from <http://www.tqsource.org/publications/October2009Brief.pdf>.
- Wenger, E.C., R. McDermott, and W.M. Snyder. 2002. *Cultivating Communities of Practice*. Cambridge, MA: Harvard Business School Press.